

# 基于轻量化网络与嵌入式的 分心行为协同检测系统

李少凡<sup>1,2</sup>, 高尚兵<sup>1,2</sup>, 张莹莹<sup>1,2</sup>, 黄 想<sup>1</sup>, 杨苏强<sup>1</sup>, 郭筱宇<sup>1</sup>

(1. 淮阴工学院计算机与软件工程学院, 江苏 淮安 223001)

(2. 淮阴工学院江苏省物联网移动互联技术工程实验室, 江苏 淮安 223001)

**[摘要]** 分心驾驶是交通事故发生的主要原因之一. 针对目前分心驾驶检测手段单一、检测种类少、检测效率低的问题, 提出一种基于轻量化网络与嵌入式的分心行为协同检测系统. 首先, 结合 Ghost 模块和通道注意力机制提出一种轻量化目标检测网络 YOLO-Ghost, 采用 CSPGBottleneck 构建 GhostDarknet 作为主干网络, 同时构建一种具有多尺度注意力机制的多特征融合模块 SE-FPN 来进行特征融合, 根据固有检测场景进行检测头优化, 以 CIOU (complete-IOU) 作为损失函数. 采用 YOLO-Ghost 识别和定位局部特征, 提出 APJ (anchor position judge) 对手动分心行为进行判定; 协同检测方面, 利用 MobileNetV3 与 YOLO-Ghost 协同进行人脸关键点回归和视线估计; 最后利用检测出的多模态信息对驾驶员当前行驶状态进行联合判定. 实验结果表明, YOLO-Ghost 的准确率和检测速度优于其他主流方法. 将算法部署到嵌入式设备中, 在 NVIDIA Jetson TX1 上实现了 20FPS 的实时检测性能, 准确性和实时性均达到检测要求.

**[关键词]** 协同检测, 人物交互, 轻量级网络, 智能交通, 深度学习

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2023)01-0025-08

## Collaborative Detection System for Distraction Behavior Based on Lightweight Network and Embedded Platform

Li Shaofan<sup>1,2</sup>, Gao Shangbing<sup>1,2</sup>, Zhang Yingying<sup>1,2</sup>, Huang Xiang<sup>1</sup>, Yang Suqiang<sup>1</sup>, Guo Xiaoyu<sup>1</sup>

(1. Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an 223001, China)

(2. Laboratory for Internet of Things and Mobile Internet Technology of Jiangsu Province, Huaiyin Institute of Technology, Huai'an 223001, China)

**Abstract:** Distracted driving is the main cause of traffic accident. In order to solve the problems of fewer kinds of distracted driving detection and poor detection efficiency, a collaborative detection system for distraction behavior based on the lightweight network and embedded platform is proposed. First of all, a lightweight object detection network YOLO-Ghost is proposed by combining Ghost module and channel attention mechanism, the CSPGBottleneck is proposed to build GhostDarknet as the backbone network, and a multi-feature fusion module SE-FPN with a multi-scale attention mechanism is proposed for feature fusion. A more comprehensive CIOU (complete-IOU) function is considered as the loss function. YOLO-Ghost is used to identify and locate local features, and APJ (anchor position judge) is proposed to judge manual distraction behavior. Secondly, MobileNetV3 and YOLO-Ghost are used to perform face key point regression and gaze estimation. Finally, the detected multimodal information is used to jointly determine the current driving state of the driver. The experimental results show that the YOLO-Ghost achieves the higher accuracy and speed than other main stream methods. At the same time, when the algorithm is deployed to the embedded device, it obtains 20FPS real-time detection performance on the NVIDIA Jetson TX1 and the accuracy and real-time performance reach the detection requirements.

**Key words:** collaborative detection, human object interaction, lightweight network, intelligent transportation, deep learning

收稿日期: 2022-09-15.

基金项目: 国家重点研发计划项目 (2018YFB1004904)、国家自然科学基金面上项目 (62076107)、江苏省高校自然科学研究重大项目 (18KJA520001)、江苏省产学研合作项目 (BY2022334)、淮阴工学院研究生科技创新计划项目 (HGYK202216).

通讯作者: 高尚兵, 博士, 教授, 研究方向: 机器学习、计算机视觉、模式识别和数据挖掘. E-mail: luxiaofen\_2002@126.com

分心驾驶是交通事故发生的主要原因之一. 据交通管理局的数据显示, 每年全国大约发生 200 万起交通事故, 其中超过 80% 是由分心驾驶造成的. 近年来, 高级驾驶辅助系统(ADAS)被许多汽车厂商采用. ADAS 使用一系列传感器来收集车内外的数据, 并对车内外的静态和动态目标进行检测和识别. 其中, 驾驶行为监控是 ADAS 的关键技术. 驾驶员分心行为检测在计算机视觉和自动驾驶领域值得深入研究.

尽管驾驶员分心行为识别是人类动作识别的一个通用问题, 但由于车内狭小复杂的环境和驾驶员分心时体态的细微差异, 目前用于人体动作识别的方法<sup>[1-3]</sup>并不适用于驾驶员分心识别.

现有的对驾驶员进行分心行为检测的方法主要包括基于传统的手工特征的方法和基于深度卷积神经网络的方法. Guo 等<sup>[4]</sup>利用颜色和形状信息对驾驶行为进行检测; Yan 等<sup>[5]</sup>联合运动历史图和金字塔式梯度方向直方图进行驾驶员的行为识别; Sharma 等<sup>[6]</sup>在多个尺度上进行密集的灰度 SIFT 特征采样, 模拟每个图片局域对于分类的贡献, 然后采用 SVM 分类器对带有权重的热力图进行分类. Koesdwiady 等<sup>[7]</sup>利用深度卷积神经网络(VGGNET)对驾驶员进行分心检测; Hu 等<sup>[8]</sup>使用一个多流 CNN 来提取不同尺度的特征融合并进行分心行为检测; Ou 等<sup>[9]</sup>对神经网络进行迁移学习来改善分心驾驶行为识别的效果; Baheti 等<sup>[10]</sup>提出了一种改进的 VGG16 结构用于驾驶员分心行为检测, 利用多种正则化技术来改善检测效果, 并研究了 Dropout、L2 正则化和归一化技术对检测结果的影响.

上述方法多采用整张图片单一视角的全局特征来进行分心行为检测, 易受车内真实驾驶场景复杂多变的影响, 无法捕捉到驾驶员发生分心行为时微小的体态变化. Le 等<sup>[11]</sup>利用 Faster-RCNN 对驾驶员的手、方向盘和手机进行检测, 并判断手是否在方向盘上以及驾驶员是否在打电话, 但由于其数据集分开标注, 需重复进行检测, 计算量大、检测种类少、视角单一, 难以应用于实际场景.

本文提出一种基于轻量化网络与嵌入式的分心行为协同检测系统. 首先, 构建一个真实驾驶场景的数据集, 并进行统一标注, 针对场景中的局部特征利用 APJ 对手动分心行为进行识别, 协同进行视线偏移检测和疲劳驾驶检测. 其次, 考虑到车载系统场景的硬件低功耗限制, 设计了一种单步检测器, 使用 CSPGBottleneck 用于组构 GhostDarknet 作为主干网络进行特征提取, 实现快速准确地检测. 最后, 以多尺度注意力金字塔作为特征融合层加强融合效果, 以考虑长宽比的 CIOU 作为损失函数. 系统结构如图 1 所示.

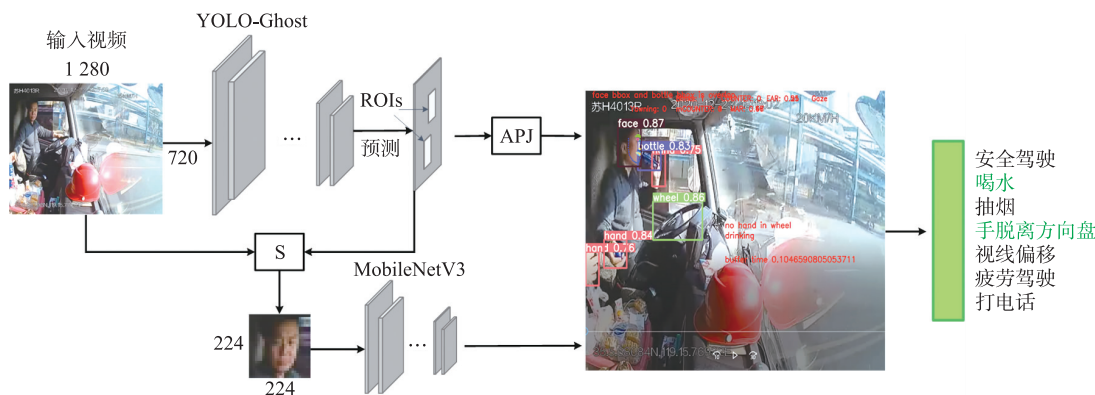


图 1 系统结构图

Fig. 1 System structure diagram

## 1 轻量化网络设计

### 1.1 YOLO 目标检测

以 R-CNN 为代表的二阶段检测器精度高, 但检测速度较慢. 以 YOLO<sup>[12]</sup>、SSD 等为首的一阶段检测器速度较快, 但精度略差. YOLOv1 对检测目标同时进行分类回归, 大幅提升了检测速度, 对目标检测的实际落地起到了巨大的促进作用. YOLOv5 可以根据使用场景的不同灵活地切换所需要的模型, 其 CSPDarkNet 主干特征提取网络采用 CSPDarkNet 网络, 结构上借鉴 CSPNet, Neck 部分采用 FPN<sup>[13]</sup> (金字塔网络) 加 PANet<sup>[14]</sup> (路径聚合网络) 的结构进行多特征融合, 同时大量使用基于 CutMix<sup>[15]</sup> 的随机 Mosaic 进行数据增强, 回归框损失函数采用 GIOU Loss<sup>[16]</sup>.

### 1.2 YOLO-Ghost 轻量级目标检测网络

为实现目标检测器的轻量化, 本文提出一个轻量级的单步检测器. 采用 CSPGBottleneck 构建 GhostDarknet

作为主干网络,以具有多尺度注意力机制的多特征融合模块 SE-FPN 作为特征融合层,根据本系统所要检测实际目标大小对原检测层进行裁剪,并通过 K-means 算法对 anchor 进行重新聚类,以获取适合检测场景的 anchor. 最后以 CIOU 作为损失函数. 整体结构如图 2 所示.

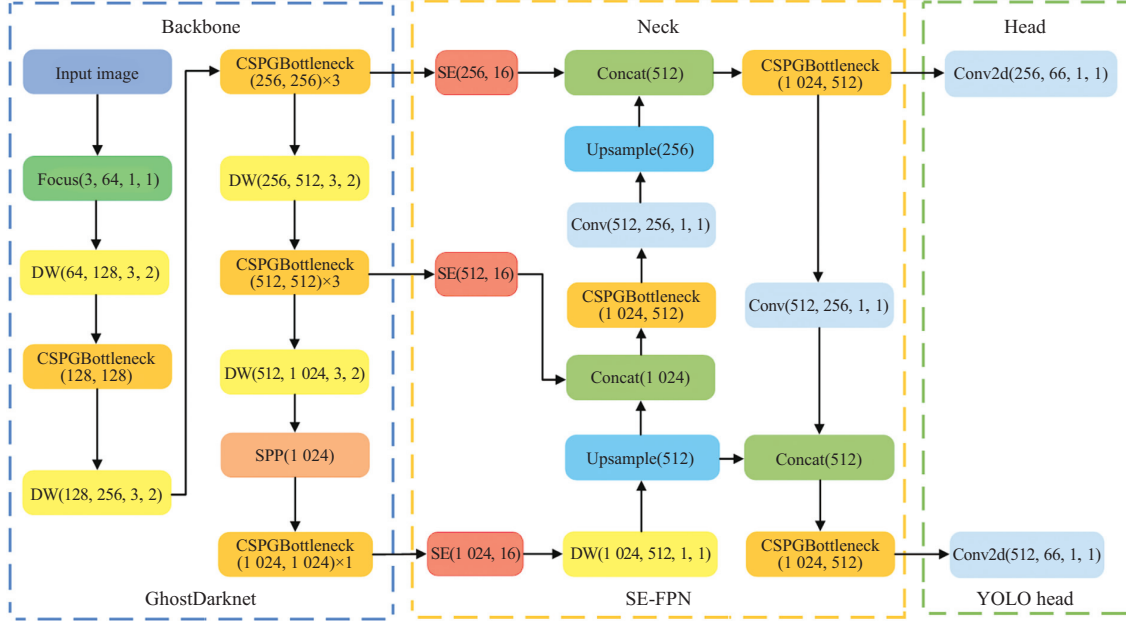


图 2 YOLO-Ghost 网络结构

Fig. 2 The structure of YOLO-Ghost

### 1.2.1 GhostDarknet

深度卷积神经网络由大量卷积模块组成,往往会带来大量的计算成本. 2020 年华为提出了一种新颖的网络 GhostNet<sup>[17]</sup>,但卷积操作会生成许多相似的特征图,导致存在大量的冗余特征图. 为避免不必要的高计算量,本文提出一种 Ghost 模块,如图 3 所示. Ghost 模块先进行卷积操作生成少量的特征图,然后对这些特征图进行线性变换生成相似的冗余特征图,得到相当识别准确率的同时避免了复杂的卷积计算,提高了检测速度.

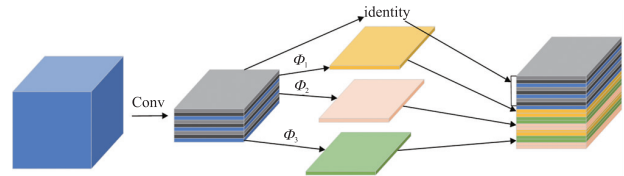


图 3 Ghost 模块

Fig. 3 Ghost module

假设输入特征图为  $h \times w \times C_{in}$ , 输出特征图为  $h' \times w' \times C_{out}$ , 把输入特征层分为  $n$  份, 卷积核尺寸为  $k$ , 则常规卷积计算量为:

$$h' \times w' \times C_{out} \times k \times k \times C_{in}. \quad (1)$$

Ghost 模块计算量为:

$$h' \times w' \times \frac{C_{out}}{n} \times k \times k \times (n-1) + h' \times w' \times \frac{C_{out}}{n} \times k \times k \times C_{in}. \quad (2)$$

从式(2)可以看出, Ghost 模块由常规卷积和深度卷积两部分组成,模型压缩率为  $n$ ,从而减少了模型计算量.

为了减少网络模型的计算量,适应硬件资源有限的车载系统,本文以 Darknet53 网络为基础结构,将原 CSPBottleneck 中的 ResNet 残差结构替换成 GhostBottleneck,形成新的 CSPGBottleneck,以此为基础重新组构 GhostDarknet 作为主干特征提取网络,将主干网络中的普通卷积替换成深度可分离卷积以减少参数数量和计算量. GhostBottleneck 和 CSPGBottleneck 网络结构如图 4 所示.

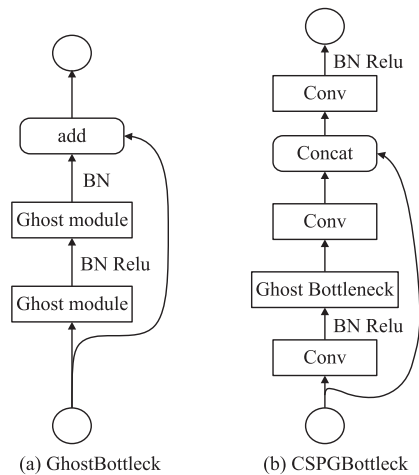


图 4 GhostBottleneck 和 CSPGBottleneck 结构图

Fig. 4 Structure of GhostBottleneck and CSPGBottleneck

### 1.2.2 SE-FPN

为了加强模型特征提取能力,本文提出了一种多尺度注意力金字塔结构 SE-FPN. 如图 5 所示,通道注意力机制 SE(squeeze-and-excitation)<sup>[18]</sup> 利用一个和通道数相同的一维向量去学习各个通道的重要性,从而增强有效的重要特征,抑制无效或不重要特征,有效增强了模型的特征提取能力. 将注意力机制对主干网络的 3 个尺度特征抽取进行特征增强,形成具有注意力机制的金字塔网络作为特征融合结构,即 SE-FPN. 该结构可以加强多特征融合<sup>[19]</sup> 的效果,提升模型的精度.

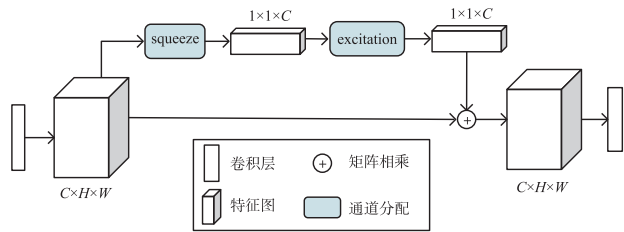


图 5 通道注意力机制结构

Fig. 5 Structure of squeeze-and-excitation

### 1.2.3 YOLO head

由于只需对驾驶场景内特定的物体(人脸、水杯、手机、香烟、方向盘、手)进行检测,目标检测头采用分类和回归同时进行的 YOLO 检测层. 经统计,这几类物体像素区间在  $15 \times 15 \sim 90 \times 90$  之间,因此采用两个 YOLO 检测层,并根据实际检测目标尺寸的大小对 anchor box 进行重新聚类,预选 anchor box 的大小为  $(22, 17), (25, 57), (37, 40), (53, 66), (77, 48), (79, 78)$ . 原网络每次前向传播需要预测的回归框数量为  $(13 \times 13 + 26 \times 26 + 52 \times 52) \times 3$ ,改进后每次前向传播需要预测的回归框数量为  $(13 \times 13 + 26 \times 26) \times 3$ ,计算量大大减少.

### 1.2.4 损失函数

YOLO-Ghost 采用 CIOU 损失函数,全面考虑了矩形的长宽比、边界框和目标之间的距离、重叠率以及尺度等各方面因素. CIOU 可进一步提升模型性能,加快损失函数收敛,增强区域定位. CIOU 的计算式为:

$$\text{CIOU} = \text{IOU} - \frac{\rho^2(b, b^{\text{gt}})}{c} - \alpha v, \quad (3)$$

式中, CIOU 为完全交并比; IOU 为交并比;  $\rho$  为两个点的欧式距离;  $b, b^{\text{gt}}$  分别代表预测框和真实框的中心点;  $c$  为能够同时包含预测框和真实框的最小闭包区域的对角线距离;  $\alpha$  为正权衡参数;  $v$  为权重函数,用以度量长宽比:

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2, \quad (4)$$

式中,  $w, w^{\text{gt}}$  分别为预测框宽和真实框宽;  $h, h^{\text{gt}}$  分别为预测框高和真实框高.

损失函数公式为:

$$\text{LOSS}_{\text{CIOU}} = 1 - \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v. \quad (5)$$

## 1.3 人脸关键点检测和视线估计

人脸关键点检测和视线估计在驾驶员行为检测中起着至关重要的作用,分别用来对驾驶员进行疲劳检测和计算驾驶员的视线偏移,这两者都是驾驶员状态监控的重要参数. 为了进一步提升检测效率,首先从 YOLO-Ghost 模型推理得到所有检测框的 4 个坐标信息,再从中找到脸部的 4 个坐标,将其从原图上裁剪下来,并进行  $S(Y(x), x)$  采样操作,对人脸进行裁剪、缩放、标准化和透视变换,然后送入轻量级网络 Mobilenetv3,并对网络结构进行微调以适应人脸关键点检测和视线估计任务.

### 1.4 交互判别方法

为了实现对驾驶员行为的准确判别,提出一种 APJ(anchor position judge)行为判别方法,利用人和物体的 Bounding Box 的位置交互关系对驾驶行为进行识别. 具体算法流程如下:

输入 模型推理得到  $\text{interference}[]$  ( $\text{interference}[]$  代表所有预测类别的 Bounding Box).

输出 当前行为类别  $\text{action}[]$ .

步骤 1 将  $\text{interference}[]$  中的所有 Bounding Box 按类别和坐标进行建类,依次加入  $\text{anchorlist}[]$  中;

步骤 2 对  $\text{anchorlist}[]$  搜索找出脸部框所在位置并保存,设定 IOU 阈值;

步骤 3 取出  $\text{anchorlist}[]$  中的 Bounding Box 进行比较,和脸部框进行 IOU 计算,若 IOU 阈值大于给



定阈值,则判定成立,并将当前行为类别加入到 `action[ ]` 中,同时移除当前的 Bounding Box;

**步骤 4** 重复步骤 3,直至 `anchorlist[ ]` 为空.

判断手脱离方向盘的步骤与此一致.

## 2 实验结果及分析

### 2.1 实验平台和参数

本文实验的模型训练在一块 Tesla V100 (16GB) 上完成;模型推理测试的设备为一台 8G 内存、NVIDIA-GTX 950M 的计算机,使用的深度学习框架为 PyTorch1.8.0.

模型训练采用随机梯度下降算法对网络模型进行更新优化,图片统一缩放到  $512 * 512$ ,批训练样本大小设置为 64,训练 epochs 设置为 200.

### 2.2 实验数据集

本文使用的分心驾驶实验数据来自真实场景下某公交平台的内部监控驾驶员行驶记录视频,数据大小为 430 GB,像素为  $1280 * 720$ ,包含了同一场景下红外摄像头拍摄的灰度图和高清摄像头拍摄的日间行车视频. 利用标注工具 LabelLmg 对其中的人脸、手部、水杯、香烟、方向盘等物体进行标注,将标注完成的图片制作成 VOC 格式的数据集,共包含 15 000 张训练集和 6 000 张验证集.

### 2.3 性能评价指标

本文实验将模型大小、参数量、平均精度 (mAP@0.5)、FLOPs 及 FPS 作为各模型的评估指标. mAP 为所有类别的平均准确率 (average precision, AP). 平均准确率定义为:

$$AP = \int_0^1 P(R) dR, \quad (6)$$

式中,  $P$  为准确率;  $R$  为召回率. 检测速度 FPS 定义为每秒处理图像的帧数.

GFLOPs 代表网络的计算复杂度 (每秒百万次浮点计算数, Giga Flops).

### 2.4 实验结果分析

#### 2.4.1 目标检测实验结果分析

将制作完成的 VOC 格式数据集送入 YOLO-Ghost 进行训练,经 200 次迭代训练,得到训练好的模型. 读取待检测视频的每一帧图像,通过训练好的网络模型进行目标检测,得到要检测的目标种类和置信度. 图 6 所示为日间目标检测效果图,图 7 所示为夜间灰度图像目标检测效果图.



图 6 日间行车目标检测效果图

Fig. 6 Object detection result of daytime driving



图 7 夜间灰度图像目标检测效果图

Fig. 7 Object detection result in gray-image at night driving

可以看出,即使在较为复杂的车内环境和不同的视角中,YOLO-Ghost 均能准确检测出驾驶员的身体部位和车内关键物品. 相比于传统方法利用整张图片的全局信息进行图片单分类,本文方法能够在对单张图片检测出多个分心行为的同时,在一定程度上减少环境中冗余信息的影响,能够更好地应对复杂环境检测.

表 1 为各目标检测模型在分心驾驶数据集上的测试效果. YOLO-Ghost 与最新的 YOLOv5s 相比,模型大小减少了 11.2M,计算复杂度下降了 12.1GFLOPs,实时性 (FPS) 提升了 22 帧/s. 可见,本文所提网络模

型更加轻量化,更适合部署于车载边缘设备中.

表 1 不同目标检测方法在分心驾驶数据集上的结果

Table 1 The results of different object detection methods on the distracted driving dataset

| 检测方法                         | 模型大小/MB | 参数量    | mAP  | GFLOPs | FPS  |
|------------------------------|---------|--------|------|--------|------|
| RetinaNet <sup>[20]</sup>    | 140.3   | 40M    | 89.4 | 61.2   | 3.2  |
| Faster-R-CNN <sup>[21]</sup> | 108.2   | 28.48M | 85.6 | 124.8  | 1.6  |
| SSD <sup>[22]</sup>          | 100.3   | 26.28M | 87.3 | 31.4   | 10.3 |
| Centernet <sup>[23]</sup>    | 124.61  | 32.67M | 86.7 | 13.0   | 6.5  |
| YOLOv5-S <sup>[24]</sup>     | 14.4    | 7.4M   | 90.1 | 17.0   | 22.1 |
| YOLO-Ghost                   | 3.2     | 1.5M   | 90.6 | 4.9    | 46.2 |

2.4.2 分心驾驶识别实验分析

选取  $t1-t21$  段真实环境下驾驶员行车监控分心视频进行检测,每段视频长 11 s,采集帧数为 25 帧/s,检测吸烟、打电话、喝水、双手脱离方向盘等分心驾驶行为. 其中,视线偏移、眨眼和打哈欠为新定义的分类. 新增分心类别的判定规则为:当预测的视线角度的 pitch(垂直方向)角或 yaw(水平方向)角大于  $40^{\circ}$  时,则判定为视线偏移;眨眼和打哈欠的判定方式通过计算眼部和嘴部的 ear(纵横比),当眼部的 ear 大于 0.2 时判定为眨眼,当嘴部的 ear 大于 0.5 时判定为打哈欠. 通过统计对比检测出的分心驾驶行为的帧数与实际分心驾驶行为的帧数计算准确率,检测结果如表 2 所示.

表 2 检测结果

Table 2 The detection result

| 行为类别    | 测试视频            | 准确率/% | FPS/(帧/s) | 行为类别 | 测试视频            | 准确率/% | FPS/(帧/s) |
|---------|-----------------|-------|-----------|------|-----------------|-------|-----------|
| 吸烟      | $t1, t2, t3$    | 95.26 | 43.12     | 视线偏移 | $t13, t14, t15$ | 93.26 | 44.36     |
| 打电话     | $t4, t5, t6$    | 94.28 | 42.26     | 眨眼   | $t16, t17, t18$ | 96.52 | 45.83     |
| 喝水      | $t7, t8, t9$    | 92.15 | 46.53     | 打哈欠  | $t19, t20, t21$ | 94.32 | 43.25     |
| 双手脱离方向盘 | $t10, t11, t12$ | 97.23 | 47.52     |      |                 |       |           |

从实验结果可以看出,分心协同检测系统对玩手机、喝水的检出率均表现在 90% 以上. 最终检测帧率均在 42FPS 以上,满足监控视频实时监测需求. 日间和夜间视角检测效果如图 8、图 9 所示. 图 8 可同时检测出喝水和手不在方向盘两种分心行为. 针对不同的视角和不同的分心行为,协同检测方法均能有效检测.



图 8 日间行车协同检测效果图

Fig. 8 Collaborative detection result of daytime driving



图 9 灰度图像协同检测效果图

Fig. 9 Collaborative detection result of gray-image

2.5 嵌入式平台实现

本文采用的嵌入式设备是 Jetson TX1 开发板,首先通过 ubuntu16.04 主机将 Jetpack 及所需组件一并烧录进 Jetson TX1 中,其中关于深度学习的组件主要有 CUDA10.2、Cudnn8.0、opencv4.2.3、Tensor7.1. 算法采用 Python 语言开发,PyTorch 作为深度学习框架,并将相机集成到编程环境中. 真实场景部署情况如图 10 所示.

分心检测系统算法流程如下:

**步骤 1** 利用 USB 摄像头采集到的图片序列,对图片进行裁剪、缩放、对比度增强等中值滤波预处理操作,将图片缩放至  $512 * 512$ ,作为模型的输入;

**步骤 2** 对图片进行标准化并输入到 YOLO-Ghost 目标检测模型中,得到当前图片的所有目标信息;

**步骤 3** 利用 APJ 判别方法得到当前的驾驶员手动分心行为  $action[]$ ;

**步骤 4** 根据第一阶段检测出的锚框信息,进行  $S(Y(x), x)$  采样操作,将采样后的图片送入 MobileNetV3 模型中进行人脸关键点估计和视线估计,得到 98 个人脸关键点和人眼视线方向  $yaw$  角和  $pitch$  角;

**步骤 5** 根据人眼部和嘴部纵横比的变化来判定是否眨眼和打哈欠,并按其在一定时间内打呵欠和眨眼的频率判断是否疲劳驾驶;根据其视线方向是否超过给定阈值来判定是否视线偏移;

**步骤 6** 结合  $action[]$  和脸部检测结果判定当前驾驶员状态,记录首次检出分心驾驶行为帧时间,当连续检出时间超过 1 s,且检出帧数量超过总帧数 60%,进行预警提示。

本算法通过网络轻量化设计,在 NVIDIA Jetson TX1 上实现了 20FPS 的实时检测性能。图 11、图 12 为在本地嵌入式设备上的检测效果。



图 10 基于嵌入式的分心监测系统

Fig. 10 Distraction monitoring system based on embedded device

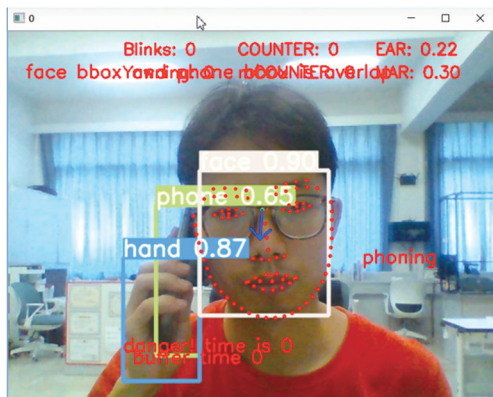


图 11 打电话检测效果

Fig. 11 Detection of phoning

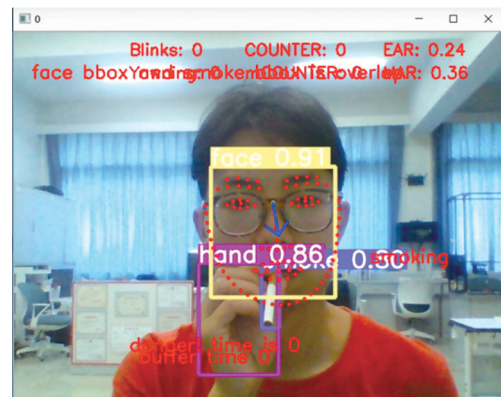


图 12 吸烟检测效果图

Fig. 12 Detection of smoking

### 3 结论

针对传统分心驾驶方法检测效率差、检测种类少的问题,提出一种基于轻量化网络与嵌入式的分心行为协同检测方法,同时对手动分心、视觉分心、精神分心进行检测,识别驾驶员分心原因,并在嵌入式设备上实现了实时检测。

尽管本文提出的分心行为协同检测方法可利用驾驶员的信息对一些常见的分心行为进行实时监控预警,但驾驶员分心受多方面影响,仅利用人的信息是不全面的,如何结合人-车-路三者对驾驶员当前状态进行融合分析是下一步要进行的工作。

### [参考文献] (References)

- [1] 王飞,胡荣林,金鹰. 基于 3D-CBAM 注意力机制的人体动作识别[J]. 南京师范大学学报(工程技术版),2021,21(1): 49-56.



- [2] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description [C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015.
- [3] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks [C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (CVPR). Santiago, Chile: IEEE, 2015.
- [4] GUO G D, LAI A. A survey on still image based human action recognition [J]. Pattern Recognition, 2014, 47(10): 3343–3361.
- [5] YAN C, COENEN F, ZHANG B L. Driving posture recognition by joint application of motion history image and pyramid histogram of oriented gradients [J]. International Journal of Vehicular Technology, 2014, 846: 719413.
- [6] SHARMA G, JURIE F, SCHMID C. Discriminative spatial saliency for image classification [C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, USA: IEEE, 2012.
- [7] KOESDWIADY A, BEDAWI S M, OU C, et al. End-to-end deep learning for driver distraction recognition [C]//Proceedings of the 2017 International Conference on Image Analysis and Recognition. Montreal, Canada: Springer, 2017.
- [8] HU Y C, LU M Q, LU X B. Driving behaviour recognition from still images by using multi-stream fusion CNN [J]. Machine Vision and Applications, 2019, 30(5): 851–865.
- [9] OU C J, OUALI C, KARRAY F. Transfer learning based strategy for improving driver distraction recognition [C]//Proceedings of the 2018 International Conference on Image Analysis and Recognition. Póvoa de Varzim, Portugal: Springer, 2018.
- [10] BAHETI B, GAJRE S, TALBAR S. Detection of distracted driver using convolutional neural network [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE, 2018.
- [11] LE T H N, ZHENG Y, ZHU C C, et al. Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas, USA: IEEE, 2016.
- [12] 梁秦嘉, 刘怀, 陆飞. 基于改进YOLOv3的运动目标分类检测算法研究 [J]. 南京师范大学学报(工程技术版), 2021, 21(4): 27–32.
- [13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017.
- [14] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018.
- [15] YUN S D, HAN D Y, CHUN S H, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019.
- [16] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019.
- [17] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020.
- [18] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(8): 2011–2023.
- [19] 殷业瑜, 高家全, 李莹. 面向印花图案检索的特征融合方法研究 [J]. 南京师大学报(自然科学版), 2022, 45(2): 118–125.
- [20] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [21] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016.
- [22] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points [EB/OL]. arXiv Preprint arXiv:1904.07850v2, 2019.
- [23] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–327.
- [24] JOCHER G. YOLOv5 [EB/OL]. [2020-08-09]. <https://github.com/ultralytics-yolov5>.

[责任编辑: 严海琳]